On the Rényi Divergence, Joint Range of Relative Entropies, and a Channel Coding Theorem

Igal Sason

Abstract

This paper starts by considering the minimization of the Rényi divergence subject to a constraint on the total variation distance. Based on the solution of this optimization problem, the exact locus of the points $\big(D(Q\|P_1), D(Q\|P_2)\big)$ is determined when P_1, P_2, Q are arbitrary probability measures which are mutually absolutely continuous, and the total variation distance between P_1 and P_2 is not below a given value. It is further shown that all the points of this convex region are attained by probability measures which are defined on a binary alphabet. This characterization yields a geometric interpretation of the minimal Chernoff information subject to a constraint on the variational distance.

This paper also derives an exponential upper bound on the performance of binary linear block codes (or code ensembles) under maximum-likelihood decoding. Its derivation relies on the Gallager bounding technique, and it reproduces the Shulman-Feder bound as a special case. The bound is expressed in terms of the Rényi divergence from the normalized distance spectrum of the code (or the average distance spectrum of the ensemble) to the binomially distributed distance spectrum of the capacity-achieving ensemble of random block codes. This exponential bound provides a quantitative measure of the degradation in performance of binary linear block codes (or code ensembles) as a function of the deviation of their distance spectra from the binomial distribution. An efficient use of this bound is considered.

Keywords: Chernoff information, distance spectrum, error exponent, maximum-likelihood decoding, relative entropy, Rényi divergence, total variation distance.

I. Introduction

The Rényi divergence, introduced in [30], has been studied so far in various information-theoretic contexts (and it has been actually used before it had a name [37]). These include generalized cutoff rates and error exponents for hypothesis testing ([1], [6], [38]), guessing moments ([2], [9]), source and channel coding error exponents ([2], [12], [22], [27], [37]), strong converse theorems for classes of networks [11], strong data processing theorems for discrete memoryless channels [28], bounds for joint source-channel coding [41], and one-shot bounds for information-theoretic problems [46].

In [14], Gilardoni derived a Pinsker-type lower bound on the Rényi divergence $D_\alpha(P\|Q)$ for $\alpha \in (0,1)$. In view of the fact that this lower bound is not tight, especially when the total variation distance $|P-Q|$ is large, this paper starts by considering the minimization of the Rényi divergence $D_\alpha(P\|Q)$, for an arbitrary $\alpha > 0$, subject to a given (or minimal) value of the total variation distance. Note that the minimization here is taken over all probability measures with a total variation distance which is not below a given value; this problem differs from the type of problems studied in [3] and [24], in connection to the minimization of the relative entropy $D(P\|Q)$ subject to a minimal value of the total variation distance with a fixed probability measure Q. The solution of this problem generalizes the problem of minimizing the relative entropy $D(P\|Q)$ subject to a given value of the total variation distance where the latter is a special case with $\alpha = 1$ (see [10], [13], [29]).

One possible way to deal with this problem stems from the fact that the Rényi divergence is a one-to-one transformation of the Hellinger divergence $\mathscr{H}_\alpha(P\|Q)$ where for $\alpha \in (0,1) \cup (1,\infty)$:

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \, \log \left(1 + (\alpha - 1) \, \mathscr{H}_\alpha(P\|Q)\right) \tag{1}$$

and $\mathscr{H}_\alpha(P\|Q)$ is an f-divergence; since the total variation distance is also an f-divergence, this problem can be viewed as a minimization of an f-divergence subject to a constraint on another f-divergence. The numerical optimization of an f-divergence subject to simultaneous constraints on f_i-divergences $(i = 1, \ldots, L)$ was recently studied in [15], where it has been shown that it suffices to restrict attention to alphabets of cardinality $L + 2$. In fact, as shown in [44, (22)], a binary alphabet suffices if there is a single constraint (i.e., $L = 1$) which is on the total variation distance. In view of (1), the same conclusion also holds when minimizing the Rényi divergence subject to a constraint on the total variation distance. To set notation, the divergences $D(P\|Q)$, $|P - Q|$, $\mathscr{H}_\alpha(P\|Q)$, $D_\alpha(P\|Q)$ are defined at the end of this section,

being consistent with the notation in [35] and [45].

This paper treats this minimization problem of the Rényi divergence in a different way. We first generalize the analysis in [10], which was used for the minimization of the relative entropy subject to a constraint on the variational distance, for proving that it suffices to restrict attention to probability measures which are defined on a binary alphabet. Furthermore, the continuation of the analysis in this paper relies on the Lagrange duality, and a solution of the Karush-Kuhn-Tucker (KKT) equations while asserting strong duality for the studied problem. The use of Lagrange duality further simplifies the computational task of the studied minimization problem.

As complementary results to the minimization problem studied in this paper, the reader is referred to [35, Section 8] which provides upper bounds on the Rényi divergence $D_\alpha(P\|Q)$ for an arbitrary $\alpha \in (0, \infty)$ as a function of either the total variation distance or relative entropy in case that the relative information is bounded.

The solution of the minimization problem of the Rényi divergence, subject to a constraint on the total variation distance, provides an elegant way for the characterization of the exact locus of the points $\big(D(Q\|P_1), D(Q\|P_2)\big)$ where P_1 and P_2 are probability measures whose total variation distance is not below a given value ε, and Q is an arbitrary probability measure. It is further shown in this paper that all the points of this convex region can be attained by a triple of probability measures (P_1, P_2, Q) which are defined on a binary alphabet.

In view of the characterization of the exact locus of these points, a geometric interpretation is provided in this paper for the minimal Chernoff information between P_1 and P_2, denoted by $C(P_1, P_2)$, subject to an ε-separation constraint on the variational distance between P_1 and P_2. It is demonstrated in the following that the intersection point at the boundary of the locus of $\big(D(Q\|P_1), D(Q\|P_2)\big)$ and the straight line $D(Q\|P_1) = D(Q\|P_2)$ is the point whose coordinates are equal to the minimal value of $C(P_1, P_2)$ under the constraint $|P_1 - P_2| \geq \varepsilon$. The reader is referred to [48], which relies on the closed-form expression in [31, Proposition 2] for the minimization of the constrained Chernoff information, and which analyzes the problem of channel-code detection by a third-party receiver via the likelihood ratio test. In the latter problem, a third-party receiver has to detect the channel code used by the transmitter by observing a large number of noise-affected codewords; this setup has applications in security or cognitive radios, or in link adaptation in some wireless technologies.

Since the Rényi divergence $D_\alpha(P\|Q)$ forms a generalization of the relative entropy $D(P\|Q)$, where the latter corresponds to $\alpha = 1$, the approach suggested in this paper for the characteri-

zation of the exact locus of pairs of relative entropies in view of a solution to a minimization problem of the Rényi divergence is analogous to the usefulness of complex analysis in solving real-valued problems. We consider the analysis of the considered problem as mathematically pleasing in its own right. Note, however, that an operational meaning of a special point at the boundary of this locus has an operational meaning in view of [48] (see the previous paragraph). The studied problem considered here differs from the study in [17] which considered the joint range of f-divergences for pairs (rather than triplets) of probability measures.

The performance analysis of linear codes under maximum-likelihood (ML) decoding is of interest for studying the potential performance of these codes under optimal decoding, and for the evaluation of the degradation in performance that is incurred by the use of sub-optimal and practical decoding algorithms. The reader is referred to [32] which is focused on this topic.

The second part of this paper derives an exponential upper bound on the performance of ML decoded binary linear block codes (or code ensembles). Its derivation relies on the Gallager bounding technique (see [32, Chapter 4], [36]), and it reproduces the Shulman-Feder bound [40] as a special case. The new exponential bound derived in this paper is expressed in terms of the Rényi divergence from the normalized distance spectrum of the code (or average distance spectrum of the ensemble) to the binomial distribution which characterizes the average distance spectrum of the capacity-achieving ensemble of fully random block codes. This exponential bound provides a quantitative measure of the degradation in performance of binary linear block codes (or code ensembles) as a function of the deviation of their (average) distance spectra from the binomial distribution, and its use is exemplified for an ensemble of turbo-block codes.

This paper is structured as follows: Section II solves the minimization problem for the Rényi divergence under a constraint on the total variation distance, Section III uses the solution of this minimization problem to obtain an exact characterization of the joint range of the relative entropies in the considered setting above. Section IV provides a new exponential upper bound on the block error probability of ML decoded binary linear block codes, which is expressed in terms of the Rényi divergence, suggests an efficient way to apply the bound to the performance evaluation of binary linear block codes (or code ensembles), and exemplifies its use. Throughout this paper, logarithms are to the base e.

We end this section by introducing the definitions and notation used in this work, which are consistent with [35], [45], and are included here for the convenience of the reader.

Definitions and Notation

We assume throughout that the probability measures P and Q are defined on a common measurable space $(\mathcal{A}, \mathscr{F})$, and $P \ll Q$ denotes that P is *absolutely continuous* with respect to Q, namely there is no event $\mathcal{F} \in \mathscr{F}$ such that $P(\mathcal{F}) > 0 = Q(\mathcal{F})$. Let $\frac{\mathrm{d}P}{\mathrm{d}Q}$ denote the Radon-Nikodym derivative (or density) of P with respect to Q.

Definition 1 (Relative entropy): The relative entropy is given by

$$D(P\|Q) = \int_{\mathcal{A}} \mathrm{d}P \, \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right). \tag{2}$$

Definition 2 (Total variation distance): The total variation distance is given by

$$|P - Q| = \int_{\mathcal{A}} \left|\frac{\mathrm{d}P}{\mathrm{d}Q} - 1\right| \mathrm{d}Q. \tag{3}$$

Definition 3 (Hellinger divergence): The Hellinger divergence of order $\alpha \in (0,1) \cup (1,\infty)$ is given by

$$\mathscr{H}_{\alpha}(P\|Q) = \frac{1}{\alpha - 1} \left(\int_{\mathcal{A}} \mathrm{d}Q \left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)^{\alpha} - 1\right). \tag{4}$$

The analytic extension of $\mathscr{H}_{\alpha}(P\|Q)$ at $\alpha = 1$ yields $\mathscr{H}_1(P\|Q) = D(P\|Q)$ (nats).

Definition 4 (Rényi divergence): The *Rényi divergence* of order $\alpha \geq 0$ is given as follows:

- If $\alpha \in (0,1) \cup (1,\infty)$, then

$$D_{\alpha}(P\|Q) = \frac{1}{\alpha - 1} \, \log\left(\int_{\mathcal{A}} \mathrm{d}Q \left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)^{\alpha}\right). \tag{5}$$

- If $\alpha = 0$, then

$$D_0(P\|Q) = \max_{\mathcal{F} \in \mathscr{F} : \, P(\mathcal{F})=1} \log\left(\frac{1}{Q(\mathcal{F})}\right). \tag{6}$$

- $D_1(P\|Q) = D(P\|Q)$ which is the analytic extension of $D_{\alpha}(P\|Q)$ at $\alpha = 1$.
- If $\alpha = +\infty$ then

$$D_{\infty}(P\|Q) = \log\left(\operatorname{ess\,sup} \frac{\mathrm{d}P}{\mathrm{d}Q}(Y)\right) \tag{7}$$

with $Y \sim Q$.

Definition 5 (Chernoff information): The Chernoff information between probability measures P_1 and P_2 is expressed as follows in terms of the Rényi divergence:

$$C(P_1, P_2) = \max_{\alpha \in [0,1]} \left\{(1 - \alpha)D_{\alpha}(P_1\|P_2)\right\} \tag{8}$$

and it is the best achievable exponent in the Bayesian probability of error for binary hypothesis testing (see, e.g., [5, Theorem 11.9.1]).

II. Minimization of the Rényi Divergence with a Constrained Total Variation Distance

In this section, we derive a tight lower bound on the Rényi divergence $D_\alpha(P_1\|P_2)$ subject to an equality constraint on the total variation distance $|P_1 - P_2| = \varepsilon$ where $\varepsilon \in [0,2)$ is fixed; alternatively, it can regarded as a minimization problem under the inequality constraint $|P_1 - P_2| \geq \varepsilon$. It is first shown that this lower bound is attained for probability measures defined on a binary alphabet, and Lagrange duality is used to further simplify the computational task of this bound. The special case where $\alpha = 1$, which is specialized to the minimization of the relative entropy subject to a fixed total variation distance, has been studied extensively, and three equivalent forms of the solution to this optimization problem were derived in [10], [13], [29].

In [14, Corollaries 6 and 9], Gilardoni derived two Pinsker-type lower bounds on the Rényi divergence of order $\alpha \in (0,1)$, expressed in terms of the total variation distance. Among these two bounds, the improved lower bound is given (in nats) by

$$D_\alpha(P\|Q) \geq \tfrac{1}{2}\,\alpha\varepsilon^2 + \tfrac{1}{9}\,\alpha(1 + 5\alpha - 5\alpha^2)\varepsilon^4, \quad \forall\,\alpha \in (0,1) \tag{9}$$

where $|P - Q| = \varepsilon$ denotes the total variation distance between P and Q. Note that in the limit where ε tend to 2 (from below), this lower bound converges to a finite value which is at most $\frac{22}{9}$; it is, however, an artifact of the lower bound in view of the next lemma.

Lemma 1:

$$\lim_{\varepsilon\uparrow 2}\ \inf_{P,Q\,:\,|P-Q|=\varepsilon} D_\alpha(P\|Q) = \infty, \quad \forall\,\alpha > 0. \tag{10}$$

Proof: See Appendix I-A. ∎

In the following, we derive a tight lower bound which is shown to be achievable by a restriction of the probability measures to a binary alphabet. For $\alpha > 0$, let

$$g_\alpha(\varepsilon) \triangleq \min_{P_1,P_2\,:\,|P_1-P_2|=\varepsilon} D_\alpha(P_1\|P_2), \tag{11}$$

$$= \min_{P_1,P_2\,:\,|P_1-P_2|\geq\varepsilon} D_\alpha(P_1\|P_2), \quad \forall\,\varepsilon \in [0,2). \tag{12}$$

In the following, we evaluate the function g_α. In view of [10, Section 2] which characterizes the minimum of the relative entropy in terms of the total variation distance, we first extend the argument in [10] to prove the next lemma.

Lemma 2: For an arbitrary $\alpha > 0$, the minimization in (11) is attained by probability measures which are defined on a binary alphabet.

Proof: See Appendix I-B. ■

The following proposition enables to calculate g_α for an arbitrary positive α.

Proposition 1: Let $\alpha \in (0,1) \cup (1,\infty)$ and $\varepsilon \in [0,2)$. The function g_α in (11) satisfies

$$g_\alpha(\varepsilon) = \min_{p,q\in[0,1]\colon |p-q|\geq\frac{\varepsilon}{2}} d_\alpha(p\|q) \tag{13}$$

where

$$d_\alpha(p\|q) \triangleq \frac{\log\left(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\right)}{\alpha-1} \tag{14}$$

denotes the binary Rényi divergence.

Proof: This directly follows from Lemma 2. ■

Proposition 2:

$$g_{\frac{1}{2}}(\varepsilon) = -\log\left(1 - \tfrac{1}{4}\,\varepsilon^2\right), \quad \forall\, \varepsilon \in [0,2) \tag{15}$$

and

$$g_2(\varepsilon) = \begin{cases} \log(1+\varepsilon^2), & \text{if } \varepsilon \in [0,1], \\ -\log\left(1 - \tfrac{1}{2}\,\varepsilon\right), & \text{if } \varepsilon \in (1,2). \end{cases} \tag{16}$$

Furthermore, for $\alpha \in (0,1)$ and $\varepsilon \in [0,2)$,

$$g_\alpha(\varepsilon) = \left(\frac{\alpha}{1-\alpha}\right) g_{1-\alpha}(\varepsilon), \tag{17}$$

and

$$g_\alpha(\varepsilon) \geq c_1(\alpha)\,\log\left(\frac{1}{1-\frac{1}{2}\varepsilon}\right) + c_2(\alpha), \tag{18}$$

where

$$c_1(\alpha) \triangleq \min\left\{1, \frac{\alpha}{1-\alpha}\right\}, \quad c_2(\alpha) \triangleq -\frac{\log 2}{1-\alpha}. \tag{19}$$

Proof: See Appendix II. ■

Remark 1: The lower bound on $g_\alpha(\cdot)$ in (18) provides another proof of Lemma 1 since it first yields that $\lim_{\varepsilon\uparrow 2} g_\alpha(\varepsilon) = \infty$ for $\alpha \in (0,1)$; this lemma also holds for $\alpha \geq 1$ since $D_\alpha(P\|Q)$ is monotonically increasing in its order α.

In the following, we use Lagrange duality to obtain an alternative expression as a solution of the minimization problem for g_α. Recall that Proposition 1 applies to every $\alpha > 0$. The following enables to simplify considerably the computational task in calculating g_α, for $\alpha \in (0,1)$.

Lemma 3: Let $\alpha \in (0,1)$ and $\varepsilon' \in (0,1)$. The function

$$f_{\alpha,\varepsilon'}(q) \triangleq \frac{\left(1 - \frac{\varepsilon'}{1-q}\right)^{\alpha-1} - \left(1 + \frac{\varepsilon'}{q}\right)^{\alpha-1}}{\left(1 + \frac{\varepsilon'}{q}\right)^{\alpha} - \left(1 - \frac{\varepsilon'}{1-q}\right)^{\alpha}}, \quad \forall q \in (0, 1-\varepsilon') \tag{20}$$

is strictly monotonically increasing, positive, continuous, and

$$\lim_{q \to 0^+} f_{\alpha,\varepsilon'}(q) = 0, \qquad \lim_{q \to (1-\varepsilon')^-} f_{\alpha,\varepsilon'}(q) = +\infty. \tag{21}$$

Proof: See Appendix III. ■

Corollary 1: For $\alpha \in (0,1)$ and $\varepsilon' \in (0,1)$, the equation

$$f_{\alpha,\varepsilon'}(q) = \tfrac{1-\alpha}{\alpha} \tag{22}$$

has a unique solution $q \in (0, 1-\varepsilon')$.

Proof: It follows from Lemma 3, and the mean value theorem for continuous functions.

■

Remark 2: Since $f_{\alpha,\varepsilon'} : (0, 1-\varepsilon') \to (0, \infty)$ is strictly monotonically increasing (see Lemma 3), the numerical calculation of the unique solution of equation (22) is easy.

An alternative simplified form for the optimization problem in Proposition 1 is next provided for orders $\alpha \in (0,1)$. Hence, Proposition 1 applies to every $\alpha > 0$, whereas the following is restricted to $\alpha \in (0,1)$. This, however, proves to be very useful in the next section in terms of obtaining a significant reduction in the computational complexity of $g_\alpha(\cdot)$ where only $\alpha \in (0,1)$ is of interest there.[1]

Proposition 3: Let $\alpha \in (0,1)$, $\varepsilon \in (0,2)$, and let $\varepsilon' = \frac{\varepsilon}{2}$. A solution of the minimization problem for $g_\alpha(\varepsilon)$ in Proposition 1 is obtained by calculating the binary Rényi divergence $d_\alpha(p\|q)$ in (14) while taking the unique solution $q \in (0, 1-\varepsilon')$ of (22), and setting $p = q + \varepsilon'$.

Proof: See Appendix IV. ■

In view of Proposition 3, the plots in Figures 1 and 2 provide numerical results.

---

[1]This saving in the computational complexity accelerated the running time of the numerical calculations in our computer by two orders of magnitude.
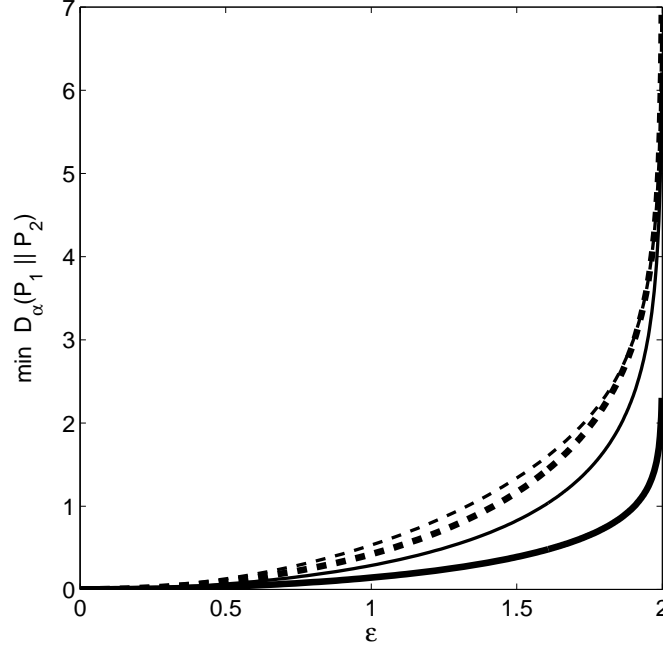
Fig. 1. A plot of the minimum of the Rényi divergence $D_\alpha(P_1\|P_2)$ subject to the constraint $|P_1 - P_2| \geq \varepsilon$ where $\varepsilon \in [0, 2)$. The curves in this plot correspond to $\alpha = 0.25$ (thick solid curve), $\alpha = 0.50$ (thin solid curve), $\alpha = 0.75$ (thick dashed curve), and $\alpha = 1.00$ (thin dashed curve, referring to the relative entropy).

III. THE LOCUS OF $(D(Q\|P_1), D(Q\|P_2))$ WITH A CONSTRAINED TOTAL VARIATION DISTANCE

In this section, we address the following question:

Question 1: What is the locus of the points $\big(D(Q\|P_1), D(Q\|P_2)\big)$ if P_1, P_2, Q are arbitrary probability measures which are mutually absolutely continuous, and $|P_1 - P_2| \geq \varepsilon$ for a given value $\varepsilon \in (0, 2)$? (none of the three probability measures is fixed).

The present section provides an exact characterization of this locus in view of the solution to the minimization problem in Section II, and the following lemma:

Lemma 4: Let P_1, P_2, Q be pairwise mutually absolutely continuous probability measures defined on a measurable space $(\mathcal{A}, \mathscr{F})$. Then, for $\alpha \in (0, 1) \cup (1, \infty)$,

$$D_\alpha(P_1\|P_2) = D(Q\|P_2) + \tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_1) + \tfrac{1}{\alpha-1} \cdot D(Q\|Q_\alpha) \tag{23}$$
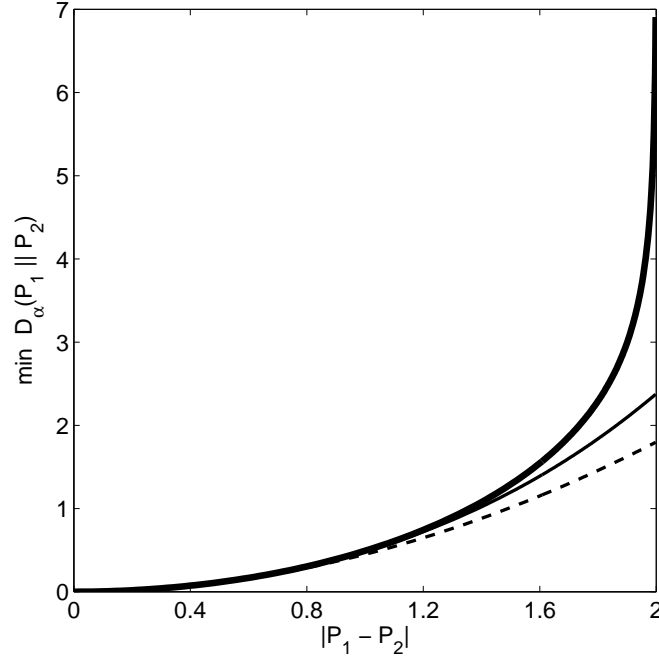
Fig. 2.   A plot of the minimum of the Rényi divergence $D_\alpha(P_1\|P_2)$ of order $\alpha = 0.90$ subject to the constraint $|P_1 - P_2| \geq \varepsilon \in [0, 2)$. The exact minimum (thick solid curve) is compared with the Pinsker-type lower bound in [14, Corollary 9] (the thin solid curve), and its weaker version in [14, Corollary 6] (the dashed curve).

where the probability measure Q_α is given by

$$\frac{\mathrm{d}Q_\alpha}{\mathrm{d}Q}(x) = \frac{\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(x)\right)^\alpha \left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}(x)\right)^{1-\alpha}}{\int_{\mathcal{A}}\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(u)\right)^\alpha \left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}(u)\right)^{1-\alpha} \mathrm{d}Q(u)}, \quad \forall\, x \in \mathcal{A}. \tag{24}$$

Proof: See Appendix V.   ■

As a corollary of Lemma 4, the following tight inequality holds, which is attributed to van Erven [7, Lemma 6.6] and Shayevitz [39, Section IV.B.8]). It will be useful for the continuation of this section, jointly with the results of Section II.

Corollary 2: Let $P_1 \lll P_2$ be mutually absolutely continuous discrete probability measures defined on a common set $\mathcal{A}$. If $\alpha \in (0, 1)$ then

$$\tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_1) + D(Q\|P_2) \geq D_\alpha(P_1\|P_2) \tag{25}$$

with equality if and only if, for every $x \in \mathcal{A}$,

$$Q(x) = \frac{P_1(x)^\alpha \, P_2(x)^{1-\alpha}}{\sum_{u \in \mathcal{A}} P_1(u)^\alpha \, P_2(u)^{1-\alpha}}. \tag{26}$$

For $\alpha > 1$, inequality (25) is reversed with the same necessary and sufficient condition for an equality.

Remark 3: The knowledge of the maximizing probability measure in (26) is required for the characterization of the exact locus which is studied in this section.

The exact locus of the points $\big(D(Q\|P_1), D(Q\|P_2)\big)$ is determined as follows: let $|P_1 - P_2| \geq \varepsilon$ for a fixed $\varepsilon \in (0, 2)$, and let $\alpha \in (0, 1)$ be chosen arbitrarily. By the tight lower bound in Section II, we have

$$D_\alpha(P_1\|P_2) \geq g_\alpha(\varepsilon) \tag{27}$$

where g_α is expressed in (13). For $\alpha \in (0, 1)$ and for a fixed value of $\varepsilon \in (0, 2)$, let $p = p^\star$ and $q = q^\star$ in $(0, 1)$ be set to achieve the global minimum in (13) (note that, without loss of generality, one can assume that $p \geq q$ since if (p, q) achieves the minimum in (13) then also $(1 - p, 1 - q)$ achieves the same minimum). Consequently, the lower bound in (27) is attained by probability measures P_1, P_2 which are defined on a binary alphabet (see Lemma 2) with

$$\begin{aligned}
P_1(0) &= p^\star = p^\star(\alpha, \varepsilon), \quad P_1(1) = 1 - p^\star; \\
P_2(0) &= q^\star = q^\star(\alpha, \varepsilon), \quad P_2(1) = 1 - q^\star.
\end{aligned} \tag{28}$$

From Corollary 2 and (27), (28), it follows that for every $\alpha \in (0, 1)$

$$g_\alpha(\varepsilon) \leq D(Q\|P_2) + \tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_1) \tag{29}$$

where equality in (29) holds if P_1 and P_2 are the probability measures in (28) which are defined on a binary alphabet, and Q is the respective probability measure in (26) which is therefore also defined on a binary alphabet. Hence, there exists a triple of probability measures P_1, P_2, Q which are defined on a binary alphabet and satisfy (29) with equality, and these probability measures are easy to calculate for every $\alpha \in (0, 1)$ and $\varepsilon \in (0, 2)$.

Remark 4: Similarly to (29), since $|P_1 - P_2| = |P_2 - P_1|$, it follows from (29) that

$$g_\alpha(\varepsilon) \leq D(Q\|P_1) + \tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_2). \tag{30}$$

By multiplying both sides of (30) by $\frac{1-\alpha}{\alpha}$ and relying on the skew-symmetry property in (17), it follows that (30) is equivalent to

$$g_{1-\alpha}(\varepsilon) \leq D(Q\|P_2) + \tfrac{1-\alpha}{\alpha} \cdot D(Q\|P_1)$$

which is (29) when $\alpha \in (0,1)$ is replaced by $1-\alpha$. Hence, since (29) holds for every $\alpha \in (0,1)$, there is no additional information in (30).

Theorem 1: The exact locus of $\big(D(Q\|P_1), D(Q\|P_2)\big)$ in the setting of Question 1 is the convex region whose boundary is the convex envelope of all the straight lines

$$D(Q\|P_2) + \tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_1) = g_\alpha(\varepsilon), \quad \forall \alpha \in (0,1) \tag{31}$$

(i.e., the boundary is the pointwise maximum of the set of straight lines in (31) for $\alpha \in (0,1)$). Furthermore, all the points in this convex region, including its boundary, are attained by probability measures P_1, P_2, Q which are defined on a binary alphabet.

Proof: Let P_1, P_2, Q be arbitrary probability measures which are mutually absolutely continuous and satisfy the ε separation condition for P_1 and P_2 in total variation. In view of Corollary 2 and since by definition $D_\alpha(P_1\|P_2) \geq g_\alpha(\varepsilon)$, it follows that the point $\big(D(Q\|P_1), D(Q\|P_2)\big)$ satisfies

$$D(Q\|P_2) + \tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_1) \geq g_\alpha(\varepsilon) \tag{32}$$

for every $\alpha \in (0,1)$; this implies that every such a point is either on or above the convex envelope of the parameterized straight lines in (31).

We next prove that a point which is below the convex envelope of the lines in (31) cannot be achieved under the constraint $|P_1 - P_2| \geq \varepsilon$. The reason for this claim is because for such a point $\big(D(Q\|P_1), D(Q\|P_2)\big)$, there is some $\alpha \in (0,1)$ for which

$$D(Q\|P_2) + \tfrac{\alpha}{1-\alpha} \cdot D(Q\|P_1) < g_\alpha(\varepsilon) \tag{33}$$

Since under the ε separation condition for P_1 and P_2 in total variation distance, $D_\alpha(P_1\|P_2) \geq g_\alpha(\varepsilon)$, then for such $\alpha \in (0,1)$, inequality (25) is violated; in view of Corollary 2, this yields that the point is not achievable under the constraint $|P_1 - P_2| \geq \varepsilon$. As an interim conclusion, it follows that the exact locus of the achievable points is the set of all points in the plane $\big(D(Q\|P_1), D(Q\|P_2)\big)$ which are on or above the convex envelope of the parameterized straight lines in (31) for $\alpha \in (0,1)$.

The next step aims to show that an arbitrary point which is located at the boundary of this region can be obtained by a triplet of probability measures $(P_1^\star, P_2^\star, Q^\star)$ which are defined on a binary alphabet, and satisfy $|P_1^\star - P_2^\star| = \varepsilon$. To that end, note that every point which is on the boundary of this region is a tangent point to one of the straight lines in (31) for some $\alpha \in (0,1)$.

Accordingly, the proper probability measures $P_1^\star$, $P_2^\star$ and Q^* can be determined as follows for a given $\varepsilon \in (0, 2)$:

a) Find the slope $s < 0$ of the tangent line at the selected point on the boundary; in view of (31), $s = -\frac{\alpha}{1-\alpha}$ yields $\alpha = -\frac{s}{1-s} \in (0, 1)$.

b) In view of Proposition 3, determine $p_1^\star, p_2^\star \in (0, 1)$ such that $|p_1^\star - p_2^\star| = \frac{\varepsilon}{2}$ and $d_\alpha(p_1^\star \| p_2^\star) = g_\alpha(\varepsilon)$. Consequently, let $P_1^\star$ and $P_2^\star$ be the probability measures which are defined on the binary alphabet with $P_1^\star(0) = p_1^\star$ and $P_2^\star(0) = p_2^\star$.

c) The respective probability measure $Q^\star = Q_\alpha^\star$ is calculated from (26), and it is therefore also defined on the binary alphabet.

Finally, we show that every interior point in the achievable region can be attained as well by a proper selection of $P_1^\star$, $P_2^\star$ and $Q^\star$ which are defined on a binary alphabet. To that end, note that every such interior point is located at the boundary of the locus of $\big(D(Q\|P_1), D(Q\|P_2)\big)$ under the constraint $|P_1 - P_2| \geq \bar\varepsilon$ with some $\bar\varepsilon \in (\varepsilon, 2)$; this follows from the fact that $g_\alpha(\cdot)$ is a strictly monotonically increasing and continuous function in $(0, 2)$, which tends to infinity as we let ε tend to 2 (see Lemma 1). It therefore follows that the suitable triplet of probability measures $(P_1^\star, P_2^\star, Q^\star)$ can be obtained by the same algorithm used for points on the boundary of this region, except for replacing ε by the larger value $\bar\varepsilon$.

This concludes the proof by first characterizing the exact locus of points, and then demonstrating that every point in this convex region (including its boundary) is attained by probability measures which are defined on the binary alphabet; the proof is also constructive in the sense of providing an algorithm to calculate such probability measures $P_1^\star, P_2^\star, Q^\star$ for an arbitrary point in this closed and convex region. ∎

As it is shown in Figure 4, the boundaries of these regions become less curvy as $\varepsilon \uparrow 2$.

A Geometric Interpretation of the Minimal Chernoff Information with a Constraint on the Variational Distance

Consider the point in Figure 4 which, in the plane of $\big(D(Q\|P_1), D(Q\|P_2)\big)$, is the intersection of the straight line $D(Q\|P_1) = D(Q\|P_2)$ and the boundary of the convex region which is characterized in Theorem 1 for an arbitrary $\varepsilon \in (0, 2)$.

In view of the proof of Theorem 1, this intersection point satisfies $D(Q_\alpha\|P_1) = D(Q_\alpha\|P_2)$ for some $\alpha \in (0, 1)$, for P_1, P_2 which are probability measures defined on a binary alphabet with $|P_1 - P_2| = \varepsilon$, and Q_α is given in (26). The equal coordinates of this intersection point
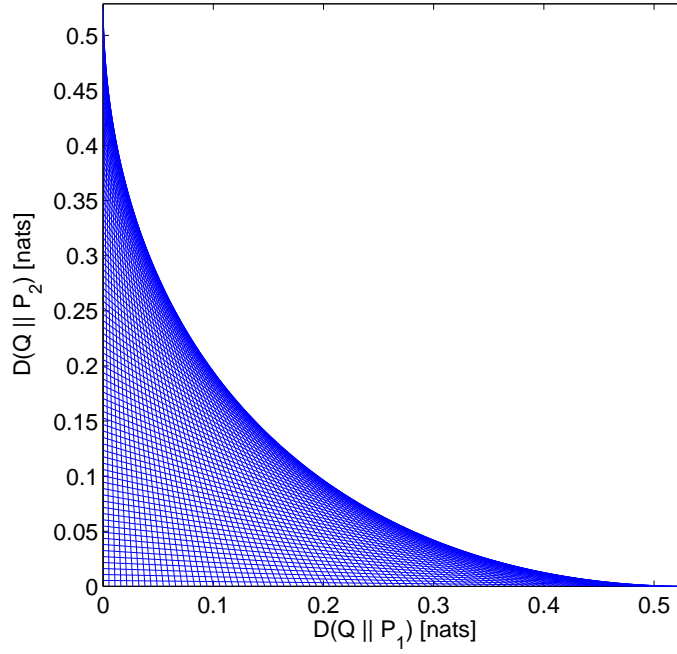
Fig. 3. The exact locus of $(D(Q\|P_1), D(Q\|P_2))$ where P_1, P_2 are arbitrary probability measures with $|P_1 - P_2| \geq 1$ with $\varepsilon = 1$. The exact locus of these relative entropies includes all the points on and above the convex envelope of the straight lines in (31), which is the convex and closed region painted in white.

are therefore equal to the Chernoff information $C(P_1, P_2)$ (see [5, Section 11.9]). Due to the symmetry of this region with respect to the straight line $D(Q\|P_1) = D(Q\|P_2)$ (this follows from the symmetry property $|P_1 - P_2| = |P_2 - P_1|$), the slope of the tangent line to the boundary of the convex region at this intersection point is $s = -1$ (see Figure 4). This yields that $\alpha = -\frac{s}{1-s} = \frac{1}{2}$, and from Proposition 2, $g_\alpha(\varepsilon) = -\log\left(1 - \frac{1}{4}\varepsilon^2\right)$. Hence, from (31) with $\alpha = \frac{1}{2}$, the equal coordinates of this intersection point are given by

$$D(Q\|P_1) = D(Q\|P_2) = -\tfrac{1}{2}\log\left(1 - \frac{1}{4}\varepsilon^2\right). \tag{34}$$

Based on [31, Proposition 2], this value is equal to the minimum of the Chernoff information subject to an ε separation constraints for P_1 and P_2 in total variation distance. We next calculate the probability measures $P_1^\star$, $P_2^\star$ and $Q^\star$ which attain this intersection point. Eq. (13) with $\alpha = \frac{1}{2}$ yields

$$-2\log\left(\sqrt{pq} + \sqrt{(1-p)(1-q)}\right) = -\log\left(1 - \tfrac{1}{4}\varepsilon^2\right) \tag{35}$$
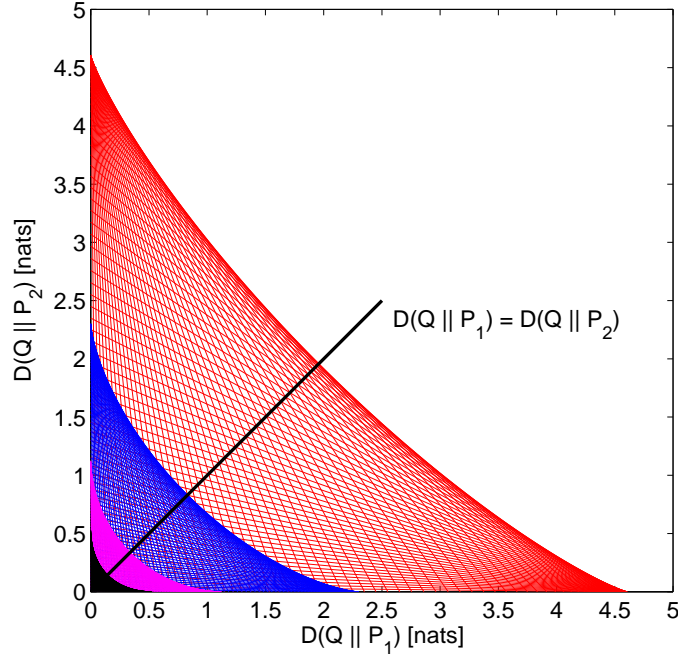
Fig. 4. This plot shows the 4 exact loci of $(D(Q\|P_1), D(Q\|P_2))$ where P_1, P_2 are arbitrary probability measures such that $|P_1 - P_2| \geq \varepsilon$, with $\varepsilon = 1.00, 1.40, 1.80, 1.98$, and $Q \ll P_1, P_2$ is an arbitrary probability measure. The exact locus which is above the convex envelope for the respective value of ε (painted in white) shrinks as the value of ε is increased, especially when ε is close (from below) to 2. The intersection of the boundary of the exact locus, for a given $\varepsilon \in [0, 2)$, with the straight line $D(Q\|P_1) = D(Q\|P_2)$ (passing through the origin) is at the point $\left(-\frac{1}{2}\log(1 - \varepsilon^2), -\frac{1}{2}\log(1 - \varepsilon^2)\right)$; the equal coordinates of this point are the minimum of the Chernoff information subject to a given total variation distance ε.

such that $p, q \in [0, 1]$ and $|p - q| = \frac{\varepsilon}{2}$. A possible solution of this equation is $p = \frac{2+\varepsilon}{4}$ and $q = \frac{2-\varepsilon}{4}$, so the respective probability measures $P_1^\star, P_2^\star$ which are defined on the binary alphabet satisfy $P_1^\star(0) = \frac{2+\varepsilon}{4}$ and $P_2^\star(0) = \frac{2-\varepsilon}{4}$; consequently, from (26), $Q(0) = Q(1) = \frac{1}{2}$ is the equiprobable distribution on the binary alphabet.

As a byproduct of the characterization of the convex region in Theorem 1, it follows that the straight line $D(Q\|P_1) = D(Q\|P_2)$ (in the plane of Figure 4) intersects the boundary of the convex region which is specified in Theorem 1 at the point whose coordinates are equal to the minimized Chernoff information subject to the constraint $|P_1 - P_2| \geq \varepsilon$. The equal coordinates of each of the 4 intersection points in Figure 4, which refer to $\varepsilon = 1.00, 1.40, 1.80, 1.98$, are equal to $-\frac{1}{2}\log\left(1 - \frac{1}{4}\varepsilon^2\right) = 0.144, 0.337, 0.830, 1.959$ nats, respectively.

IV. A Performance Bound for Coded Communications via the Rényi Divergence

A. New Exponential Upper Bound

This section derives an exponential upper bound on the performance of binary linear block codes, expressed in terms of the Rényi divergence. Similarly to [19], [20], [21], [23], [25], [33, Section 3.B], [36], [40] and [43], the upper bound in the next theorem quantifies the degradation in the performance of block codes under ML decoding in terms of the deviation of their distance spectra from the binomially distributed (average) distance spectrum of the capacity-achieving ensemble of random block codes.

Theorem 2: Consider a binary linear block code of length N and rate $R = \frac{\log(M)}{N}$ where M designates the number of codewords. Let $S_0 = 0$ and, for $l \in \{1, \ldots, N\}$, let S_l be the number of non-zero codewords of Hamming weight l. Assume that the transmission of the code takes place over a memoryless, binary-input and output-symmetric channel. Then, the block error probability under ML decoding satisfies

$$P_{\mathrm{e}} = P_{\mathrm{e}|0} < \exp\left(-N \sup_{r \geq 1} \max_{0 \leq \rho' \leq \frac{1}{r}} \left[E_0\left(\rho', \underline{q} = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right) - \rho'\left(rR + \frac{D_s(P_N \| Q_N)}{N}\right)\right]\right) \quad (36)$$

where $s \triangleq s(r) = \frac{r}{r-1}$ for $r \geq 1$ (with the convention that $s = \infty$ for $r = 1$), Q_N is the binomial distribution with parameter $\frac{1}{2}$ and N independent trials (i.e., $Q_N(l) = 2^{-N}\binom{N}{l}$ for $l \in \{0, 1, \ldots, N\}$), P_N is the PMF defined by $P_N(l) = \frac{S_l}{M-1}$ for $l \in \{0, \ldots, N\}$, $D_s(\cdot \| \cdot)$ is the Rényi divergence of order s (i.e., $D_s(P \| Q) = \frac{1}{s-1} \log\left(\sum_x P(x)^s Q(x)^{1-s}\right)$ where $s > 1$ here), and $E_0(\rho, \underline{q})$ designates the Gallager random coding error exponent in [12, Eq. (5.6.14)].

Before proving Theorem 2, we relate this exponential bound to previously reported bounds.

Remark 5: Note that the loosening of the bound by taking $r = 1$ and, respectively, $s = \infty$ gives the upper bound

$$\begin{aligned}
P_{\mathrm{e}} = P_{\mathrm{e}|0} &< \exp\left(-N \max_{0 \leq \rho' \leq 1}\left[E_0\left(\rho', \underline{q} = \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right) - \rho'\left(R + \frac{D_\infty(P_N \| Q_N)}{N}\right)\right]\right) \\
&\overset{\mathrm{(a)}}{=} \exp\left(-N E_{\mathrm{r}}\left(R + \frac{D_\infty(P_N \| Q_N)}{N}\right)\right) \\
&\overset{\mathrm{(b)}}{=} \exp\left(-N E_{\mathrm{r}}\left(R + \frac{1}{N} \log \max_{0 \leq l \leq N} \frac{P_N(l)}{Q_N(l)}\right)\right) \\
&\overset{\mathrm{(c)}}{=} \exp\left(-N E_{\mathrm{r}}\left(R + \frac{1}{N} \log \max_{0 \leq l \leq N} \frac{S_l}{e^{-N(\log 2 - R)}\binom{N}{l}}\right)\right)
\end{aligned}$$

which coincides with the Shulman-Feder bound [40]. Equality (a) follows from the definition of the Gallager random coding exponent $E_{\mathrm{r}}(R)$ in [12, Eq. (5.6.16)] where the symmetric input distribution $\underline{q} = \left(\frac{1}{2}, \frac{1}{2}\right)$ is the optimal input distribution for any memoryless, binary-input output-symmetric channel, equality (b) follows from the expression of the Rényi divergence of order infinity (see, e.g., [8, Theorem 6]), and equality (c) follows from the definition of the PMFs P_N and Q_N in Theorem 2.

Remark 6: The proof of Theorem 2 is based on the framework of the Gallager bounds in [32, Chapter 4] and [36]. Specifically, it has an overlap with [36, Appendix A]. Unlike the analysis in [36, Appendix A], working with the Rényi divergence of order $s \geq 1$, instead of the relative entropy as a lower bound (see [36, Eq. (A19)]) reveals a need for an optimization of the error exponent, which leads to the error exponent in Theorem 2. Namely, if the value of $r \geq 1$ is increased then the value of $s = \frac{r}{r-1} \geq 1$ is decreased, and therefore $D_s(P_N \| Q_N)$ is also decreased (unless it is zero, see [8, Theorem 3]; note that P_N and Q_N do not depend on the parameters r and s, so they stay un-affected by varying the values of these parameters). The maximization of the error exponent in Theorem 2 aims at finding a proper balance between the two summands rR and $\frac{D_s(P_N \| Q_N)}{N}$ on the right-hand side of (36), while also performing an optimization over the second dependent variable $\rho' \in \left[0, \frac{1}{r}\right]$.

We proceed now with the proof of Theorem 2.

Proof: The proof of Theorem 2 is based on the framework of the Gallager bounds in [32, Chapter 4] and [36]. Specifically, it relies on [36, Appendix A]. We explain in the following how our proof differs from the analysis in [36, Appendix A]. From [36, Eq. (A17)], we have that for every $\rho' \in \left[0, \frac{1}{r}\right]$

$$P_{\mathrm{e}|0} < M^{\rho' r} \exp\left(-N E_0\left(\rho', \underline{q} = \left(\frac{1}{2}, \frac{1}{2}\right)\right)\right) \left(\sum_{l=0}^{N} Q_N(l) \left(\frac{P_N(l)}{Q_N(l)}\right)^s\right)^{\frac{r\rho'}{s}}. \qquad (37)$$

From this point, we deviate from the analysis in [36, Appendix A]. Since $\frac{1}{r} + \frac{1}{s} = 1$ where $r, s \geq 1$, we have

$$\left(\sum_{l=0}^{N} Q_N(l) \left(\frac{P_N(l)}{Q_N(l)}\right)^s\right)^{\frac{r\rho'}{s}}$$

$$= \exp\left(\frac{r\rho'}{s} \cdot \log\left(\sum_{l=0}^{N} P_N(l)^s Q_N(l)^{1-s}\right)\right)$$

$$= \exp\left(\frac{\rho'}{s-1} \cdot \log\left(\sum_{l=0}^{N} P_N(l)^s \, Q_N(l)^{1-s} \right) \right)$$

$$= \exp\big(\rho' D_s(P_N \| Q_N) \big) \tag{38}$$

where $D_s(P_N \| Q_N)$ is the Rényi divergence of order s from P_N to Q_N. This enables to refer to the Rényi divergence of order $s \geq 1$, instead of lower bounding this quantity by the relative entropy, and consequently loosening the bound (see [36, Eq. (A19)]). Note that since the Rényi divergence is monotonically increasing in its order (see, e.g., [8, Theorem 3]) and the Rényi divergence of order 1 is particularized to the relative entropy, the inequality $D_s(P_N \| Q_N) \geq D(P_N \| Q_N)$ holds. The combination of (37) and (38) gives

$$P_{\text{e}|0} < \exp(NR\rho' r) \, \exp\left(-N \, E_0\big(\rho', \underline{q} = (\tfrac{1}{2}, \tfrac{1}{2}) \big) \right) \exp\big(\rho' D_s(P_N \| Q_N) \big)$$

$$= \exp\left(-N \left[E_0\big(\rho', \underline{q} = (\tfrac{1}{2}, \tfrac{1}{2}) \big) - \rho'\left(rR + \frac{D_s(P_N \| Q_N)}{N} \right) \right] \right), \quad 0 \leq \rho' \leq \frac{1}{r}. \tag{39}$$

A maximization of the error exponent in (39) with respect to the parameters $r \geq 1$ and $\rho' \in [0, \frac{1}{r}]$ (recall that $s = s(r) = \frac{r}{r-1} > 1$) gives the upper bound in (36). ∎

B. Application of Theorem 2

An efficient use of Theorem 2 for the performance evaluation of binary linear block codes (or coee ensembles) is suggested in the following by borrowing a concept of bounding from [23], which has been further studied, e.g., in [32], [33], [43], and combining it with the new bound in Theorem 2. In order to utilize the Shulman-Feder bound for binary linear block codes in a clever way, it has been suggested in [23] to partition the binary linear block code $\mathcal{C}$ into two subcodes $\mathcal{C}_1$ and $\mathcal{C}_2$ where $\mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{C}$ and $\mathcal{C}_1 \cap \mathcal{C}_2 = \{0\}$ is the all-zero codeword. The first subcode $\mathcal{C}_1$ contains the all-zero codeword and all the codewords of $\mathcal{C}$ whose Hamming weights l belong to a subset $\mathcal{L} \subseteq \{1, 2, ..., N\}$, while $\mathcal{C}_2$ contains the other codewords of $\mathcal{C}$ which have Hamming weights of $l \in \mathcal{L}^{\text{c}} \triangleq \{1, 2, ..., N\} \setminus \mathcal{L}$, together with the all-zero codeword. From the symmetry of the channel, $P_{\text{e}} = P_{\text{e}|0} \leq P_{\text{e}|0}(\mathcal{C}_1) + P_{\text{e}|0}(\mathcal{C}_2)$ where $P_{\text{e}|0}(\mathcal{C}_1)$ and $P_{\text{e}|0}(\mathcal{C}_2)$ designate the conditional ML decoding error probabilities of $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively, given that the all-zero codeword is transmitted. Note that although the code $\mathcal{C}$ is linear, its two subcodes $\mathcal{C}_1$ and $\mathcal{C}_2$ are in general non-linear. One can rely on different upper bounds on the conditional error probabilities $P_{\text{e}|0}(\mathcal{C}_1)$ and $P_{\text{e}|0}(\mathcal{C}_2)$, i.e., we may bound $P_{\text{e}|0}(\mathcal{C}_1)$ by invoking Theorem 2, due to its tightening of the Shulman-Feder bound (see Remark 5), and also rely on an alternative

approach for obtaining an upper bound on $P_{\mathrm{e}|0}(\mathcal{C}_2)$ (e.g., it is possible to rely on the union bound with respect to the fixed composition codes of the subcode $\mathcal{C}_2$). The idea behind this partitioning is to include in the subcode $\mathcal{C}_1$ the codewords of all the Hamming weights whose distance spectrum is close enough to the binomial distribution Q_N (see Theorem 2) in the sense that the additional term $\frac{D_s(P_N \| Q_N)}{N}$ in the exponent of (36) has a marginal effect on the conditional ML decoding error probability of the subcode $\mathcal{C}_1$.

Theorem 2 can be applied as well to *ensembles* of binary linear block codes. The verify this claim, let $\mathcal{C}$ be an ensemble of binary linear block codes. The proof of Theorem 2 follows from the Duman and Salehi bounding technique [36] which leads to the derivation of [36, Eq. (A.11)]. By taking the expectation on the RHS of [36, Eq. (A.11)] with respect to the code ensemble $\mathcal{C}$ and invoking Jensen's inequality, the same bound holds while S_l, as it is defined in Theorem 2, is replaced by the expectation $\overline{S_l} \triangleq \mathbb{E}_{\mathcal{C}}[S_l]$ with respect to the code ensemble $\mathcal{C}$. This enables to replace P_N on the RHS of (36) with $\overline{P}_N$ where

$$\overline{P}_N(l) \triangleq \frac{\mathbb{E}_{\mathcal{C}}[S_l]}{M-1}, \quad \forall l \in \{0, \dots, N\},$$

which therefore justifies the generalization of Theorem 2 to code ensembles of binary linear block codes.

As it is exemplified in Section IV-C, Theorem 2 can be efficiently applied to ensembles of turbo-like codes in the same way that it was demonstrated to be efficient in [43]. Similarly to Theorem 2, the bound in [43, Theorem 3.1] forms another refinement of the Shulman-Feder bound, and the novelty in the former bound is the obtained tightening of the Shulman-Feder bound via the use of the Rényi divergence.

C. An Example: Performance Bounds for an Ensemble of Turbo-Block Codes

We conclude this section by an example which applies this bounding technique to the ensemble of uniformly interleaved turbo codes whose two component codes are chosen uniformly at random from the ensemble of (1072, 1000) binary systematic linear block codes. The transmission of these codes takes place over an additive white Gaussian noise (AWGN) channel, and the codes are BPSK modulated and coherently detected. The calculation of the average distance spectrum of this ensemble has been performed in [43, Section 5.D], which is required for the calculation of the upper bound in (36) where the PD P_N is replaced by its expected value over the ensemble (i.e., the normalization of the average distance spectrum by the number of

codewords, as it is defined in Theorem 2). In the following, two upper bounds on the block error probability are compared under ML decoding: the first one is the tangential-sphere bound (TSB) of Herzberg and Poltyrev (see [18], [26], [32, Section 3.2.1]), and the second bound follows from the suggested combination of the union bound and Theorem 2. Note that an optimal partitioning has been performed, in a way which is conceptually similar to [43, Algorithm 1], for obtaining the tightest bound which is obtained by combining the union bound and Theorem 2.

A comparison of the two bounds shows an advantage of the latter combined bound over the TSB in a similar way to [43, upper plot of Fig. 8] (e.g., providing a gain of about 0.2 dB over the TSB for a block error probability of 10^{-3}). Note that the Shulman-Feder bound is rather loose in this case due to the significant deviation of the ensemble distance spectrum from the binomial distribution at low and high Hamming weights. Furthermore, we note that the advantage of the proposed bound over the TSB in this example is consistent with the analysis in [26] and [42], demonstrating a gap between the random coding error exponent of Gallager and the corresponding error exponents that follow from the TSB and some of its improved versions. Recall that the random coding error exponent of Gallager achieves the channel capacity, whereas the random coding error exponent that follows from the TSB (or some of its improved variants) does not achieve the capacity of a binary-input AWGN channel for BPSK modulated fully random block codes, where the gap to capacity is especially pronounced for high coding rates. In this example, the rate of the ensemble is 0.8741 bits per channel use.

APPENDIX I
PROOFS OF LEMMAS 1 AND 2

A. Proof of Lemma 1

For $\alpha = \frac{1}{2}$, $D_{\frac{1}{2}}(P\|Q) = -2\log Z(P,Q)$ where $Z(P,Q) \triangleq \sum_x \sqrt{P(x)Q(x)}$ denotes the Bhattacharyya coefficient between the two PDs P, Q. We have

$$D_{\frac{1}{2}}(P\|Q) \geq -\log\left(1 - \frac{1}{4}\varepsilon^2\right) \tag{I.1}$$

where $|P-Q| = \varepsilon$ (see, e.g., [31, Proposition 1]; inequality (I.1) is known in quantum information theory with respect to the relation between the trace distance and fidelity [47, Section 9.3]). Hence, (I.1) implies that (10) holds for $\alpha = \frac{1}{2}$. Since $D_\alpha(P\|Q)$ is monotonically increasing in its order α (see [8, Theorem 3]), it follows that (10) also holds for $\alpha \geq \frac{1}{2}$. Finally, due to the skew-symmetry property of D_α (see [8, Proposition 2]) where $D_\alpha(P\|Q) = \left(\frac{\alpha}{1-\alpha}\right) D_{1-\alpha}(Q\|P)$

for $\alpha \in (0, 1)$, and since the total variation distance is a symmetric measure and $\frac{\alpha}{1-\alpha} > 0$ for $\alpha \in (0, 1)$, the satisfiability of (10) for $\alpha \in (\frac{1}{2}, 1)$ yields that it also holds for $\alpha \in (0, \frac{1}{2})$.

B. Proof of Lemma 2

Let $P_1 \ll P_2$ be probability measures which are defined on a common measurable space $(\mathcal{A}, \mathscr{F})$. Denote by $\phi \colon \mathcal{A} \to \{1, 2\}$ the mapping given by

$$\phi(x) = \begin{cases} 1, & \text{if } \frac{\mathrm{d}P_1}{\mathrm{d}P_2}(x) \geq 1, \\ 2, & \text{if } \frac{\mathrm{d}P_1}{\mathrm{d}P_2}(x) < 1 \end{cases}$$

and let Q_i, for $i \in \{1, 2\}$, be given by

$$Q_i(j) \triangleq \int_{\{x \in \mathcal{A} \colon \phi(x) = j\}} \mathrm{d}P_i(x), \quad \forall\, i, j \in \{1, 2\}. \tag{I.2}$$

Consequently, we have

$$\begin{aligned} |P_1 - P_2| &= \int_{\mathcal{A}} \left| \frac{\mathrm{d}P_1}{\mathrm{d}P_2}(x) - 1 \right| \mathrm{d}P_2(x) \\ &= \int_{\{x \in \mathcal{A} \colon \phi(x) = 1\}} \left(\frac{\mathrm{d}P_1}{\mathrm{d}P_2}(x) - 1 \right) \mathrm{d}P_2(x) + \int_{\{x \in \mathcal{A} \colon \phi(x) = 2\}} \left(1 - \frac{\mathrm{d}P_1}{\mathrm{d}P_2}(x) \right) \mathrm{d}P_2(x) \\ &= \big(Q_1(1) - Q_2(1) \big) + \big(Q_2(2) - Q_1(2) \big) \\ &= \sum_{j \in \{1, 2\}} \big| Q_1(j) - Q_2(j) \big| \\ &= |Q_1 - Q_2|. \end{aligned} \tag{I.3}$$

From the data processing theorem for the Rényi divergence (see [8, Theorem 9]),

$$D_\alpha(P_1 \| P_2) \geq D_\alpha(Q_1 \| Q_2) \tag{I.4}$$

where Q_1 and Q_2 are the probability measures which are defined on the binary alphabet (see (I.2)). The lemma follows by combining (I.3) and (I.4).

APPENDIX II
PROOF OF PROPOSITION 2

Eq. (15) follows from the equality $D_{\frac{1}{2}}(P \| Q) = -2 \log Z(P, Q)$ where $Z(P, Q)$ is the Bhattacharyya coefficient between P, Q, and since (see [31, Proposition 1])

$$\max_{P, Q \colon |P - Q| = \varepsilon} Z(P, Q) = \sqrt{1 - \tfrac{1}{4} \varepsilon^2}, \quad \forall\, \varepsilon \in [0, 2).$$

To prove (16), note that $D_2(P_1\|P_2) = \log\big(1 + \chi^2(P_1, P_2)\big)$ where

$$\chi^2(P_1\|P_2) \triangleq \int \left(\frac{\mathrm{d}P_1}{\mathrm{d}P_2} - 1\right)^2 \mathrm{d}P_2$$

denotes the χ^2-divergence between the probability measures P_1 and P_2 (which is the Hellinger divergence of order 2). One can derive a closed-form expression for g_2 by relying on the closed-form solution of a minimization of the χ^2-divergence $\chi^2(P_1\|P_2)$ subject to the constraint $|P_1 - P_2| = \varepsilon \in [0, 2)$, which is given by (see [29, Eq. (58)])

$$\min_{P_1,P_2:\,|P_1-P_2|=\varepsilon} \chi^2(P_1\|P_2) = \begin{cases} \varepsilon^2, & \text{if } \varepsilon \in [0, 1], \\ \frac{\varepsilon}{2-\varepsilon}, & \text{if } \varepsilon \in (1, 2). \end{cases}$$

Eq. (17) follows from the skew-symmetry property of the Rényi divergence [8, Proposition 2].

The lower bound on g_α in (18) follows from (13), which implies that for $\alpha \in (0, 1)$ and $\varepsilon \in [0, 2)$

$$g_\alpha(\varepsilon) = \frac{\log\left(\max_{p,q\in[0,1]:\,|p-q|\geq\frac{\varepsilon}{2}} \big(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}\big)\right)}{\alpha - 1} \tag{II.1}$$

and, we have

$$0 \leq \max_{p,q\in[0,1]:\,|p-q|\geq\frac{\varepsilon}{2}} \big(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}\big)$$

$$\leq \max_{p,q\in[0,1]:\,|p-q|\geq\frac{\varepsilon}{2}} p^\alpha q^{1-\alpha} + \max_{p,q\in[0,1]:\,|p-q|\geq\frac{\varepsilon}{2}} (1-p)^\alpha(1-q)^{1-\alpha}$$

$$= 2 \max_{p,q\in[0,1]:\,|p-q|\geq\frac{\varepsilon}{2}} p^\alpha q^{1-\alpha}$$

$$= 2 \max\left\{\left(1 - \tfrac{1}{2}\varepsilon\right)^\alpha, \ \left(1 - \tfrac{1}{2}\varepsilon\right)^{1-\alpha}\right\}. \tag{II.2}$$

The lower bound on g_α in (18) follows from the combination of (II.1) and (II.2).

APPENDIX III

PROOF OF LEMMA 3

For $\alpha \in (0, 1)$ and $\varepsilon' \in (0, 1)$, we have

$$\lim_{q\to 0^+} \left(1 + \frac{\varepsilon'}{q}\right)^{\alpha-1} = 0, \qquad \lim_{q\to 0^+} \left(1 + \frac{\varepsilon'}{q}\right)^{\alpha} = +\infty,$$

$$\implies \lim_{q\to 0^+} f_{\alpha,\varepsilon'}(q) = \lim_{q\to 0^+} \frac{(1 - \varepsilon')^{\alpha-1}}{\left(1 + \frac{\varepsilon'}{q}\right)^{\alpha} - (1 - \varepsilon')^{\alpha}} = 0,$$

and

$$\lim_{q \to (1-\varepsilon')^-} \left(1 - \frac{\varepsilon'}{1-q}\right)^{\alpha-1} = +\infty, \qquad \lim_{q \to (1-\varepsilon')^-} \left(1 - \frac{\varepsilon'}{1-q}\right)^{\alpha} = 0,$$

$$\implies \lim_{q \to (1-\varepsilon')^-} f_{\alpha,\varepsilon'}(q) = \lim_{q \to (1-\varepsilon')^-} \frac{\left(1 - \frac{\varepsilon'}{1-q}\right)^{\alpha-1} - (1-\varepsilon')^{1-\alpha}}{(1-\varepsilon')^{-\alpha} - \left(1 - \frac{\varepsilon'}{1-q}\right)^{\alpha}} = +\infty.$$

This proves the two limits in (21).

We prove in the following that $f_{\alpha,\varepsilon'}(\cdot)$ is strictly increasing on the interval $\left[\frac{1-\varepsilon'}{2}, 1-\varepsilon'\right)$, and we also prove later in this appendix that this function is monotonically increasing on the interval $\left(0, \frac{1-\varepsilon'}{2}\right]$. These two parts of the proof yield that $f_{\alpha,\varepsilon'}(\cdot)$ is strictly monotonically increasing on the interval $(0, 1-\varepsilon')$. The positivity of $f_{\alpha,\varepsilon'}$ on $(0, 1-\varepsilon')$ follows from the first limit in (21), jointly with the monotonicity of this function which is proved in the following.

For a proof that $f_{\alpha,\varepsilon'}(\cdot)$ is strictly monotonically increasing on $\left[\frac{1-\varepsilon'}{2}, 1-\varepsilon'\right)$, this function (see (20)) is expressed as follows:

$$f_{\alpha,\varepsilon'}(q) = \left(1 + \frac{\varepsilon'}{q}\right)^{-1} u_\alpha\big(z_{\varepsilon'}(q)\big) \tag{III.1}$$

where

$$z_{\varepsilon'}(q) \triangleq \frac{1 - \frac{\varepsilon'}{1-q}}{1 + \frac{\varepsilon'}{q}}, \tag{III.2}$$

$$u_\alpha(t) \triangleq \begin{cases} \frac{t^{\alpha-1}-1}{1-t^\alpha}, & \text{if } t \in (0,1) \cup (1, \infty), \\ \frac{1-\alpha}{\alpha}, & \text{if } t = 1. \end{cases} \tag{III.3}$$

Note that u_α in (III.3) was defined to be continuous at $t = 1$. In order to proceed, we need the following two lemmas:

Lemma III.1: Let $\varepsilon' \in (0,1)$. The function $z_{\varepsilon'}$ in (III.2) is strictly monotonically increasing on $\left(0, \frac{1-\varepsilon'}{2}\right]$, and it is strictly monotonically decreasing on $\left[\frac{1-\varepsilon'}{2}, 1-\varepsilon'\right)$. This function is also positive on $(0, 1-\varepsilon')$.

Proof: $z_{\varepsilon'}(q) > 0$ for $q \in (0, 1-\varepsilon')$ since $1 - \frac{\varepsilon'}{1-q} > 0$, and $1 + \frac{\varepsilon'}{q} > 0$. In order to prove the monotonicity properties of $z_{\varepsilon'}$, note that its derivative satisfies the equality

$$\frac{\mathrm{d}}{\mathrm{d}q} z_{\varepsilon'}(q) = \varepsilon' z_{\varepsilon'}(q) \left(\frac{1}{q(\varepsilon' + q)} - \frac{1}{(1-q)(1-\varepsilon'-q)}\right) \tag{III.4}$$

which is derived by taking logarithms on both sides of (III.2), followed by their differentiation. By setting the derivative of $z_{\varepsilon'}(q)$ (with respect to q) to zero, we have $q = \frac{1-\varepsilon'}{2}$. Since $z_{\varepsilon'}(q) > 0$ for $q \in (0, 1-\varepsilon')$, it follows from (III.4) that $z'_{\varepsilon'}(q) > 0$ for $q \in \left(0, \frac{1-\varepsilon'}{2}\right)$, and $z'_{\varepsilon'}(q) < 0$ for

$q \in \left(\frac{1-\varepsilon'}{2}, 1 - \varepsilon' \right)$. Hence, $z_{\varepsilon'}$ is strictly monotonically increasing on $\left(0, \frac{1-\varepsilon'}{2} \right]$, and it is strictly monotonically decreasing on $\left[\frac{1-\varepsilon'}{2}, 1 - \varepsilon' \right)$. ∎

Lemma III.2: Let $\alpha \in (0, 1)$. The function u_α in (III.3) is strictly monotonically decreasing and positive on $(0, \infty)$.

Proof: Differentiation of u_α in (III.3) gives that for $t > 0$

$$u'_\alpha(t) = \frac{t^{\alpha-2} \left(t^\alpha - \alpha t + \alpha - 1 \right)}{(t^\alpha - 1)^2}. \tag{III.5}$$

Note that $\frac{d}{dt} \left(t^\alpha - \alpha t + \alpha - 1 \right) = \alpha(t^{\alpha-1} - 1)$, so the derivative is zero at $t = 1$, it is positive if $t \in (0, 1)$, and it is negative if $t \in (1, \infty)$. This implies that $t^\alpha - \alpha t + \alpha - 1 \leq 0$ for every $t \in (0, \infty)$, and it is satisfied with equality if and only if $t = 1$. From (III.5), it follows that u_α is strictly monotonically decreasing on $(0, \infty)$. Since $\lim_{t \to \infty} u_\alpha(t) = 0$ (see (III.3)) and u_α is strictly monotonically decreasing on $(0, \infty)$ then it is positive on this interval. ∎

From Lemmas III.1 and III.2, it follows that $z_{\varepsilon'}$ is strictly monotonically decreasing and positive on $\left[\frac{1-\varepsilon'}{2}, 1 - \varepsilon' \right)$, and u_α is strictly monotonically decreasing and positive on $(0, \infty)$. This therefore implies that the composition $u_\alpha \left(z_{\varepsilon'}(\cdot) \right)$ is strictly monotonically increasing and positive on the interval $\left[\frac{1-\varepsilon'}{2}, 1 - \varepsilon' \right)$. Hence, from (III.1), since $f_{\alpha,\varepsilon'}(\cdot)$ is expressed as a product of two positive and strictly monotonically increasing functions on $\left[\frac{1-\varepsilon'}{2}, 1 - \varepsilon' \right)$, also $f_{\alpha,\varepsilon'}$ has these properties on this interval. This completes the first part of the proof where we show that $f_{\alpha,\varepsilon'}(\cdot)$ is strictly monotonically increasing and positive on $\left[\frac{1-\varepsilon'}{2}, 1 - \varepsilon' \right)$.

We prove in the following that $f_{\alpha,\varepsilon'}(\cdot)$ is also strictly monotonically increasing and positive on $\left(0, \frac{1-\varepsilon'}{2} \right]$. For this purpose, the function $f_{\alpha,\varepsilon'}$ is expressed in the following alternative way:

$$f_{\alpha,\varepsilon'}(q) = \frac{1}{1 - \frac{\varepsilon'}{q-1}} \left(\frac{1 - \frac{\varepsilon'}{q-1}}{1 + \frac{\varepsilon'}{q}} \right)^\alpha \frac{1 - \left(\frac{1 + \frac{\varepsilon'}{q}}{1 - \frac{\varepsilon'}{q-1}} \right)^{\alpha-1}}{1 - \left(\frac{1 - \frac{\varepsilon'}{1-q}}{1 + \frac{\varepsilon'}{q}} \right)^\alpha}$$

$$= \left(1 - \frac{\varepsilon'}{1-q} \right)^{-1} r_\alpha \left(z_{\varepsilon'}(q) \right) \tag{III.6}$$

where $z_{\varepsilon'}$ is defined in (III.2), and

$$r_\alpha(t) \triangleq \begin{cases} \frac{t^\alpha(1 - t^{1-\alpha})}{1 - t^\alpha}, & \text{if } t \in (0, \infty) \setminus \{1\}, \\ \frac{1-\alpha}{\alpha}, & \text{if } t = 1. \end{cases} \tag{III.7}$$

Note that it follows from Lemma III.1 and (III.2) that

$$z_{\varepsilon'}(q) \leq z_{\varepsilon'} \left(\frac{1-\varepsilon'}{2} \right) = \left(\frac{1-\varepsilon'}{1+\varepsilon'} \right)^2 < 1$$

so the composition $r_\alpha\big(z_{\varepsilon'}(\cdot)\big)$ in (III.6) is independent of $r_\alpha(1)$; the value of $r_\alpha(1)$ is defined in (III.7) to obtain the continuity of r_α, which leads to the following lemma:

Lemma III.3: For $\alpha \in (0,1)$, the function r_α in (III.7) is strictly monotonically increasing and positive on $(0,\infty)$.

Proof: A differentiation of r_α in (III.7) gives

$$r'_\alpha(t) = \frac{(1-\alpha)t^\alpha + \alpha t^{\alpha-1} - 1}{(t^\alpha - 1)^2} \tag{III.8}$$

so the sign of r'_α is the same as of $(1-\alpha)t^\alpha + \alpha t^{\alpha-1} - 1$. Since $\alpha \in (0,1)$, and

$$\frac{d}{dt}\big((1-\alpha)t^\alpha + \alpha t^{\alpha-1} - 1\big) = \alpha(1-\alpha)t^{\alpha-2}(t-1)$$

it follows that the last derivative is negative for $t \in (0,1)$, zero at $t = 1$, and positive for $t \in (1,\infty)$. This implies that $t = 1$ is a global minimum of the numerator of r'_α (see (III.8)), so

$$(1-\alpha)t^\alpha + \alpha t^{\alpha-1} - 1 \geq 0, \quad \forall t \in (0,\infty)$$

and equality holds if and only if $t = 1$. It therefore follows from (III.8) that $r'_\alpha(t) > 0$ for $t \in (0,\infty) \backslash \{1\}$, so $r_\alpha(\cdot)$ is strictly monotonically increasing on $(0,\infty)$. Since $\lim_{t\to 0} r_\alpha(t) = 0$, the monotonicity of $r_\alpha(\cdot)$ on $(0,\infty)$ yields that it is positive on this interval. ∎

From Lemmas III.1 and III.3, $z_{\varepsilon'}$ is strictly monotonically increasing and positive on $\big(0, \frac{1-\varepsilon'}{2}\big]$, and r_α is strictly monotonically increasing and positive on $(0,\infty)$. This implies that the composition $r_\alpha\big(z_{\varepsilon'}(\cdot)\big)$ is strictly monotonically increasing and positive on the interval $\big(0, \frac{1-\varepsilon'}{2}\big]$. From (III.6), $f_{\alpha,\varepsilon'}$ is expressed as a product of two strictly increasing and positive functions on the interval $\big(0, \frac{1-\varepsilon'}{2}\big]$, which implies that $f_{\alpha,\varepsilon'}(\cdot)$ also has these properties on this interval. This completes the second part of the proof where we show that $f_{\alpha,\varepsilon'}(\cdot)$ is strictly monotonically increasing and positive on $\big(0, \frac{1-\varepsilon'}{2}\big]$. The combination of the two parts of this proof completes the proof of Lemma 3.

APPENDIX IV

PROOF OF PROPOSITION 3

The proof relies on the Lagrange duality and KKT conditions, where strong duality is first asserted by verifying the satisfiability of Slater's condition.

Let $\alpha \in (0,1)$, $\varepsilon \in (0,2)$, and $\varepsilon' = \frac{\varepsilon}{2}$. Solving (13) is equivalent to solving the optimization problem

$$\text{maximize} \quad p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}$$

$$\text{subject to} \tag{IV.1}$$

$$\begin{cases} p,q \in [0,1], \\ |p-q| \geq \varepsilon' \end{cases}$$

where p,q are the optimization variables. The objective function of the optimization problem (IV.1) is concave for $\alpha \in (0,1)$, so this maximization problem is a convex optimization problem. Since the problem is also strictly feasible at an interior point of the domain in (IV.1), Slater's condition yields that strong duality holds for this optimization problem (see [4, Section 5.2.3]). Note that the replacement of p,q with $1-p$ and $1-q$, respectively, does not affect the value of the objective function and the satisfiability of the constraints in (IV.1). Consequently, it can be assumed with loss of generality that $p \geq q$; together with the inequality constraint $|p-q| \geq \varepsilon'$, it gives that $p - q \geq \varepsilon'$. The Lagrangian of the dual problem is given by

$$L(p,q,\lambda) = p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha} + \lambda(q - p + \varepsilon')$$

and the KKT conditions lead to the following set of equations:

$$\begin{cases} \frac{\partial L}{\partial p} = \alpha \big[p^{\alpha-1} q^{1-\alpha} - (1-p)^{\alpha-1}(1-q)^{1-\alpha} \big] - \lambda = 0, \\[2mm] \frac{\partial L}{\partial q} = (1-\alpha) \big[p^\alpha q^{-\alpha} - (1-p)^\alpha (1-q)^{-\alpha} \big] + \lambda = 0, \\[2mm] \frac{\partial L}{\partial \lambda} = q - p + \varepsilon' = 0. \end{cases} \tag{IV.2}$$

Eliminating λ from the first equation in (IV.2), and substituting it into the second equation gives

$$(1-\alpha) \left[\left(\frac{p}{q} \right)^\alpha - \left(\frac{1-p}{1-q} \right)^\alpha \right] + \alpha \left[\left(\frac{p}{q} \right)^{\alpha-1} - \left(\frac{1-p}{1-q} \right)^{\alpha-1} \right] = 0. \tag{IV.3}$$

From the third equation of (IV.2), Substituting $p = q + \varepsilon'$ into (IV.3), and re-arranging terms gives the equation $f_{\alpha,\varepsilon'}(q) = \frac{1-\alpha}{\alpha}$, where $f_{\alpha,\varepsilon'}$ is the function in (20).

APPENDIX V

PROOF OF LEMMA 4

For $\alpha \in (0, \infty) \setminus \{1\}$, the following equalities hold:

$$D(Q\|P_2) + \frac{\alpha}{1-\alpha} \cdot D(Q\|P_1) + \frac{1}{\alpha-1} \cdot D(Q\|Q_\alpha)$$

$$\overset{(a)}{=} \frac{1}{\alpha-1} \int_{\mathcal{A}} \mathrm{d}Q \, \log \left(\frac{\mathrm{d}Q}{\mathrm{d}P_2} \right)^{\alpha-1} + \frac{1}{1-\alpha} \int_{\mathcal{A}} \mathrm{d}Q \, \log \left(\frac{\mathrm{d}Q}{\mathrm{d}P_1} \right)^{\alpha} + \frac{1}{\alpha-1} \int_{\mathcal{A}} \mathrm{d}Q \, \log \left(\frac{\mathrm{d}Q}{\mathrm{d}Q_\alpha} \right)$$

$$\overset{(b)}{=} \frac{1}{\alpha-1} \int_{\mathcal{A}} \mathrm{d}Q(x) \, \log \left(\left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(x) \right)^{\alpha} \left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}(x) \right)^{1-\alpha} \left(\frac{\mathrm{d}Q_\alpha}{\mathrm{d}Q}(x) \right)^{-1} \right)$$

$$\overset{(c)}{=} \frac{1}{\alpha-1} \int_{\mathcal{A}} \mathrm{d}Q(x) \, \log \left(\int_{\mathcal{A}} \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(u) \right)^{\alpha} \left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}(u) \right)^{1-\alpha} \mathrm{d}Q(u) \right)$$

$$\overset{(d)}{=} \frac{1}{\alpha-1} \, \log \left(\int_{\mathcal{A}} \left(\frac{\mathrm{d}P_1}{\mathrm{d}Q}(u) \right)^{\alpha} \left(\frac{\mathrm{d}P_2}{\mathrm{d}Q}(u) \right)^{1-\alpha} \mathrm{d}Q(u) \right)$$

$$\overset{(e)}{=} D_\alpha(P_1\|P_2)$$

where (a) follows from the equality

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \, \log \left(\int \mathrm{d}R \left(\frac{\mathrm{d}P}{\mathrm{d}R} \right)^{\alpha} \left(\frac{\mathrm{d}Q}{\mathrm{d}R} \right)^{1-\alpha} \right) \tag{V.1}$$

where R is an arbitrary probability measure such that $P, Q \ll R$; (b) holds since P_1, P_2, Q are mutually absolutely continuous which also yields that $Q \ll\gg Q_\alpha$ (in view of (24)), (c) follows from (24), (d) holds since Q is a probability measure, and (e) follows from (V.1) (recall that $Q \ll P_1, P_2$).

Acknowledgment

REFERENCES

[1] F. Alajaji, P. N. Chen and Z. Rached, "Csiszár's cutoff rates for the general hypothesis testing problem," *IEEE Trans. on Information Theory*, vol. 50, no. 4, pp. 663–678, April 2004.

[2] R. Atar and N. Merhav, "Information-theoretic applications of the logarithmic probability comparison bound," *IEEE Trans. on Information Theory*, vol. 61, no. 10, pp. 5366–5386, October 2015.

[3] D. Berend, P. Harremoës and A. Kontorovich, "Minimum KL-divergence on complements of L_1 balls," *IEEE Trans. on Information Theory*, vol. 60, no. 6, pp. 3172–3177, June 2014.

[4] S. Boyd and L. Vanderberghe, *Convex Optimization*, Cambridge University Press, 2004.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, second edition, 2006.

[6] I. Csiszár, "Generalized cutoff rates and Rényi information measures," *IEEE Trans. on Information Theory*, vol. 41, no. 1, pp. 26–34, January 1995.

[7] T. van Erven, *When Data Compression and Statistics Disagree: Two Frequentist Challenges for the Minimum Description Length Principle*, Ph.D. dissertation, Leiden university, Leiden, the Netherlands, 2010.

[8] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. on Information Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[9] T. van Erven and P. Harremoës, "Rényi divergence and majorization," *Proceedings of the 2010 IEEE International Symposium on Information Theory*, pp. 1335–1339, Austin, Texas, USA, June 2010.

[10] A. A. Fedotov, P. Harremoës and F. Topsøe, "Refinements of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 49, no. 6, pp. 1491–1498, June 2003.

[11] S. L. Fong and V. Y. F. Tan, "Strong converse theorems for classes of multimessage multicast networks: a Rényi divergence approach," July 2014. [Online]. Available: `http://arxiv.org/abs/1407.2417`.

[12] R. G. Gallager, *Information Theory and Reliable Communications*, John Wiley, 1968.

[13] G. L. Gilardoni, "On the minimum f-divergence for given total variation," *Comptes Rendus Mathematique*, vol. 343, no. 11–12, pp. 763–766, 2006.

[14] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences," *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5377–5386, November 2010.

[15] A. Guntuboyina, S. Saha and G. Schiebinger, "Sharp inequalities for f-divergences," *IEEE Trans. on Information Theory*, vol. 60, no. 1, pp. 104–121, January 2014.

[16] P. Harremoës, "Interpretations of Rényi entropies and divergences," *Physica A*, vol. 365, pp. 57–62, 2006.

[17] P. Harremoës and I. Vajda, "On pairs of f-divergences and their joint range," *IEEE Trans. on Information Theory*, vol. 57, no. 6, pp. 3230–3235, June 2011.

[18] H. Herzberg and G. Poltyrev, "The error probability of M-ary PSK block coded modulation schemes," *IEEE Trans. on Communications*, vol. 44, no. 4, pp. 427-433, April 1996.

[19] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving LDPC codes through puncturing," *IEEE Trans. on Information Theory*, vol. 54, no. 10, pp. 4698–4706, October 2008.

[20] C. H. Hsu and A. Anastasopoulos, "Capacity-achieving codes with bounded graphical complexity and maximum-likelihood decoding," *IEEE Trans. on Information Theory*, vol. 56, no. 3, pp. 992–1006, March 2010.

[21] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1451–1461, June 2002.

[22] Y. Mansour, M. Mohri and A. Rostamizadeh, "Multiple source adaptation and the Rényi divergence," *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, pp. 367–374, Montreal, Canada, June 2009.

[23] G. Miller and D. Burshtein, "Bounds on the ML decoding error probability of LDPC codes," *IEEE Trans.on Information Theory*, vol. 47, no. 7, pp. 2696–2710, November 2001.

[24] E. Ordentlich and M. J. Weinberger, "A distribution dependenet refinement of Pinsker's inequality," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1836–1840, May 2005.

[25] H. D. Pfister and P. H. Siegel, "The serial concatenation of rate-1 codes through uniform interleavers," *IEEE Trans. on Information Theory*, vol. 49, no. 6, pp. 1425–1438, June 2003.

[26] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. on Information Theory*, vol. 40, no. 4, pp. 1284–1292, July 1994.

[27] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," *Proceedings of the Forty-Eighth Annual Allerton Conference*, pp. 1327–1333, Illinois, USA, October 2010.

[28] M. Raginsky, "Logarithmic Sobolev inequalities and strong data processing theorems for discrete channels," *Proceedings of the 2013 IEEE International Symposium on Information Theory*, pp. 419–423, Istanbul, Turkey, July 2013.

[29] M. D. Reid and R. C. Williamson, "Information, divergence and risk for binary experiments," *Journal of Machine Learning Research*, vol. 12, no. 3, pp. 731–817, March 2011.

[30] A. Rényi, "On measures of entropy and information," *Proceedings of the 4th Berekely Symposium on Probability Theory and Mathematical Statistics*, pp. 547–561, Berekeley, California, USA, 1961.

[31] I. Sason, "Tight bounds for symmetric divergence measures and a refined bound for lossless source coding," *IEEE Trans. on Information Theory*, vol. 61, no. 2, pp. 701–707, February 2015.

[32] I. Sason and S. Shamai, *Performance Analysis of Linear Codes under Maximum-Likelihood Decoding: A Tutorial, Foundations and Trends in Communications and Information Theory*, vol. 3, no. 1–2, pp. 1–222, NOW Publishers, Delft, the Netherlands, July 2006.

[33] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1611–1635, July 2003.

[34] I. Sason and S. Verdú, "Upper bounds on the relative entropy and Rényi divergence as a function of total variation distance for finite alphabets," *Proceedings of the 2015 IEEE Information Theory Workshop (ITW 2015)*, pp. 214–218, Jeju Island, Korea, October 11–15, 2015.

[35] I. Sason and S. Verdú, "Bounds among f-divergences," submitted to the *IEEE Trans. on Information Theory*, August 2015. [Online]. Available at `http://arxiv.org/abs/1508.00335`.

[36] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3029–3051, December 2002.

[37] C. Shannon, R. Gallager and E. Berlekamp, "Lower bounds to error probability for decoding on discrete memoryless channels - part I," *Information and Control*, vol. 10, pp. 65–103, February 1967.

[38] O. Shayevitz, "On Rényi measures and hypothesis testing," *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pp. 800–804, Saint Petersburg, Russia, August 2011.

[39] O. Shayevitz, "A note on a characterization of Rényi measures, and its relation to composite hypothesis testing," December 2010. [Online]. Available at `http://arxiv.org/abs/1012.4401`.

[40] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes", *IEEE Trans. on Information Theory*, vol. 45, no. 6, pp. 2101–2104, September 1999.

[41] S. Tridenski, R. Zamir and A. Ingber, "The Ziv-Zakai-Rényi bound for joint source-channel coding," *IEEE Trans. on Information Theory*, vol. 61, no. 8, pp. 4293–4315, August 2015.

[42] M. Twitto and I. Sason, "On the error exponents of some improved tangential-sphere bounds," *IEEE Trans. on Information Theory*, vol. 53, no. 3, pp. 1196–1210, March 2007.

[43] M. Twitto, I. Sason and S. Shamai, "Tightened upper bounds on the ML decoding error probability of binary linear block codes," *IEEE Trans. on Information Theory*, vol. 53, no. 4, pp. 1495–1510, April 2007.

[44] I. Vajda, "On f-divergence and singularity of probability measures," *Periodica Mathematica Hungarica*, vol. 2, no. 1–4, pp. 223–234, 1972.

[45] S. Verdú, *Information Theory*, in preparation.

[46] N. A. Warsi, "One-shot bounds for various information-theoretic problems using smooth min and max Rényi divergences," *Proccedings of the 2013 IEEE Information Theory Workshop*, pp. 434–438, Seville, Spain, September 2013.

[47] M. M. Wilde, *Quantum Information Theory*, Cambridge University Press, New York, 2013.

[48] A. D. Yardi. A. Kumar, and S. Vijayakumaran, "Channel-code detection by a third-party receiver via the likelihood ratio test," *Proceedings of the 2014 IEEE International Symposium on Information Theory*, pp. 1051–1055, Honolulu, Hawaii, USA, July 2014.